The Use of 2D-LC-MS/MS in Disease Characterization and Global Proteomics

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for Graduation *With Distinction* in Chemistry from the Undergraduate College of the Ohio State University

By

Jeffrey Clark Joyner

The Ohio State University
May 2006

Project Advisor: Dr. Michael Freitas, Department of Chemistry

TABLE OF CONTENTS

ABSTRACT

Proteins occupy more than 50% of the dry weight of the average human cell. The proteins present in a given cell at a given time reflect the function and specialization of that cell. In addition to the specialization of cells within the body leading to various normal protein states, protein 'sub-states' can be induced by various conditions such as infections or cancer *[9,21]*. These disease-induced protein sub-states result in both qualitative and quantitative differences in both ribosomal production of proteins and post-translational modification of proteins *[2]*. Each disease-induced sub-state therefore has its own protein profile or signature. For diseases such as cancer, which can easily progress undetected and whose manifestations have grave consequences, statistical comparisons of a patient's protein profile with many cancer-state protein profiles can be made for early detection. Two-dimensional-liquid chromatography-tandem mass spectrometry (2D-LC-MS/MS) holds much promise in the analysis of such disease-induced protein profiles. 2D-LC-MS/MS uses two means of separation prior to online mass analysis, whereas one-dimensional (1D) methods use only one means of separation. The increased separation of a 2D system allows for much more complete mass analysis of proteins and results in the identification and characterization of many more proteins and modifications than with a 1D system *[15]*. By comparing patients' protein profiles with a comprehensive proteomic database, 2D-LC-MS/MS could serve as the critical step in an efficient early-detection method for diseases such as cancer. Before protein profiles can be accurately compared and linked with diseases, a standardized technique must be used, and mechanisms that account for experimental variations must be implemented. The 2D-LC-MS/MS technique proposed here is both very useful and can be standardized easily.

1) INTRODUCTION

556,902 Americans died from cancer in the year 2003, with deaths from cancer in 2003 constituting 22.7% of all deaths in the United States *[8]*. Cancer is the second leading cause of death in the United States and remains an illusive disease to researchers and health specialists. Although it is known that certain lifestyle habits can lead to cancer (having a diet high in fat, smoking, and not protecting oneself from harmful UV radiation are some examples), many people are understandably reluctant to change their habits, because they are either not well enough informed, cannot easily foresee the consequences of their habits, or do not feel like investing their time and energy in monitoring their diets or habits. To help our friends, there has been much work by organizations to reduce public ignorance toward the disease. Additionally, many research groups around the world have been working diligently to find biological solutions to the cancer epidemic. There have been many high-impact findings about the detection, prevention, and treatment of the disease *[4,22]*. However, what is lacking in cancer research is a method for grouping all of the information into one complete, easily understandable collection. This task calls for a bioinformatics-based solution, in which a standardized analysis technique can be combined through a comprehensive user interface with an extensive database to compare a patient's samples with other patients' samples and use this information to detect for diseases such as cancer. One analysis technique with versatile and universal functionality is the use of 2D-LC-MS/MS to analyze protein expression in humans. It has been shown that many types of cancer involve major changes in the protein expression in cells, and therefore protein profiles that correspond to different

types of cancer can be recorded. This has the potential for inexpensive, early detection of cancer and could thereby reduce the number of deaths from cancer.

2) BIOLOGICAL OVERVIEW

Before discussing procedures for 2D-LC-MS/MS that can reproducibly represent nearly all cellular proteins that are present, it is necessary to first review some fundamental facts regarding proteins' locations, production, and functions within cells as well as other important cellular processes. A typical eukaryotic cell contains, among many other things, some important locations where proteins are most abundant: inner and outer membranes, ribosomes, the nucleus, and the cytoplasm. Within the nucleus, the DNA codes for nearly everything that happens within the organism—external/environmental factors account for everything else. Specifically, certain segments of the DNA encode for the production of all proteins produced by the organism.

2.1) CELL SPECIALIZATION

Throughout the lifetime of a human, stem cells, which are highly abundant at the beginning of development, repeatedly reproduce until they eventually become divided into specialized groups of cells, such as skin cells, liver cells, blood cells, brain cells, or muscle cells *[18]*. There are on the order of 10^{14} specialized cells in an adult human *[17]*. The same DNA is present in all of the specialized cells as was present in the stem cells, with some exceptions, such as adult red blood cells. As cells become more specialized, various factors influence the protein expression by causing certain parts of their DNA to become inactive and certain parts to become active. The ways in which

DNA is packed in the nucleus plays a crucial role in determining what segments of the DNA are expressed. Histones are a class of nuclear proteins involved in the packing of DNA that are very influential in determining what DNA segments are expressed during transcription *[9,10,13,21]*. Interestingly, certain segments of the DNA determine the structure of these histones, and these histones determine what DNA segments are expressed, thereby creating a feedback loop. As the expressed segments of DNA become increasingly dependent on the cell's function, corresponding protein changes occur. It is through this process that the protein expression in a cell becomes very dependent on the function of the cell. The following figure represents the differentiation of cells from stem cells as cells are called upon to serve various functions.

FIGURE 2.1.1: Cell Differentiation in Human Tissues*

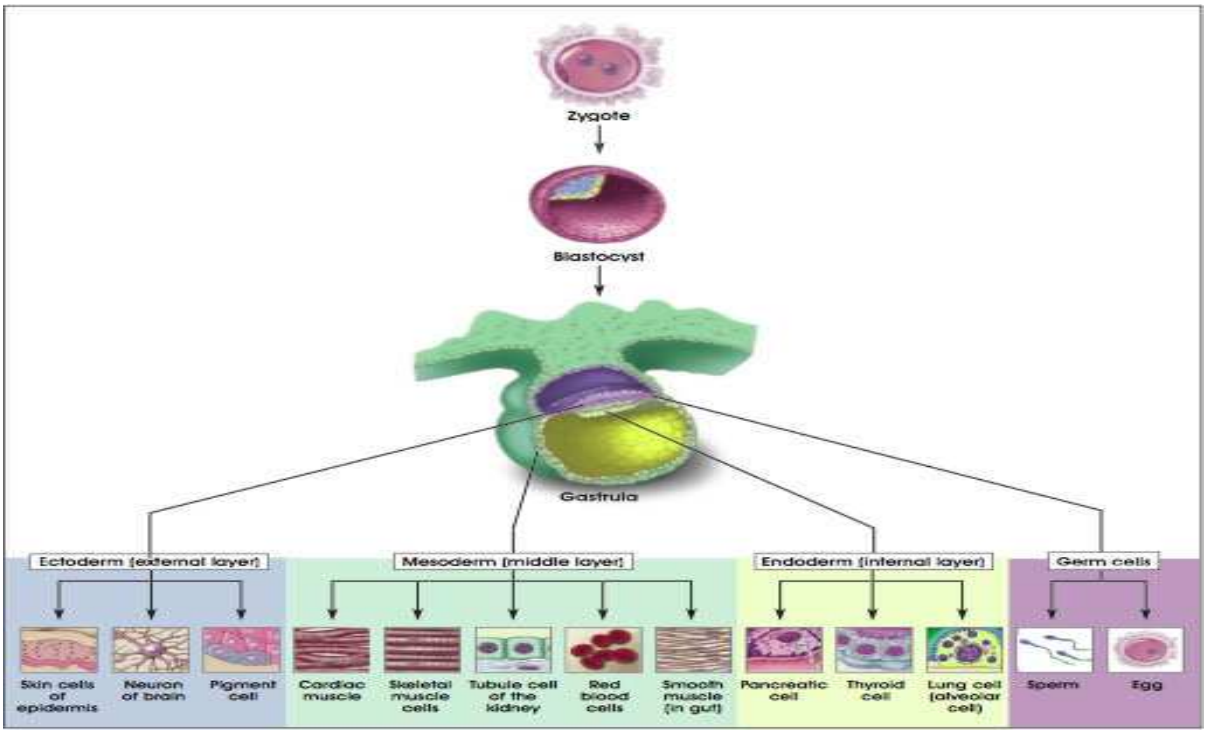


*Figure from *[17]*

DNA within the nuclei of most human cells is estimated to encode for >30,000 different proteins *[1]*. However, the proteins that exist in any given cell at a given time are highly specific and represent only a small percentage of the possible proteins encoded by the DNA. This select amount of proteins within each cell is on the order of magnitude of 10 to 10^3 distinct proteins per cell. In addition to the amount of distinct proteins in a cell being dependent on the specialization of the cell, the abundance or concentration of each type of protein present is also dependent on the cell's specialization. For example, in a red blood cell, before the nucleus is expelled, the hemoglobin encoding segments of the cell's DNA are highly expressed in the protein manufacturing process. Due to a red blood cell's high specificity, the cell is likely to have a relatively narrow distribution of distinct proteins. In an adult red blood cell, hemoglobin accounts for roughly 90% of the cell's dry weight, whereas other cells may have no single protein that accounts for more than 5% of the dry weight *[3]*. A cell in a stomach wall uses, to a larger extent than in other cells, segments of its DNA that create H^+ gradient generating membrane proteins. Some cells, such as in glands, are specialized to create proteins that are excreted into the surrounding serum.

2.2) TRANSCRIPTION AND TRANSLATION

In order to protect the DNA, cells have evolved to use RNA to copy information from DNA through the process of transcription. This allows the DNA's information to be copied without the DNA having to leave the protection of the nucleus. RNA polymerase copies the information from DNA to RNA through the mechanism shown in Figure 2.2.1. The RNA that contains the DNA's information, known as messenger RNA (mRNA),

delivers the information to the ribosomes. The ribosomes are the sites at which proteins are created within the cell. The delivery of the DNA's information in the form of mRNA to the ribosome then causes respective amino acids to be formed into a chain of amino acids. Within the mRNA, specific triples of base pairs, called codons, correspond to each amino acid used by the ribosome in the synthesis of an amino acid chain. Therefore the order of codons in the mRNA corresponds to the first order structure of the produced amino acid chain. This amino acid chain, or peptide sequence, eventually grows longer until the process is complete, at which point the chain folds into a finished protein, sometimes referred to as a polypeptide. It is through this process that the information encoded within DNA is used to make proteins; this process is called translation. The following figure depicts the transcription of mRNA from DNA and the translation of information encoded in mRNA to the formation of proteins.

FIGURE 2.2.1: Illustration of Translation of Information from DNA to Proteins*

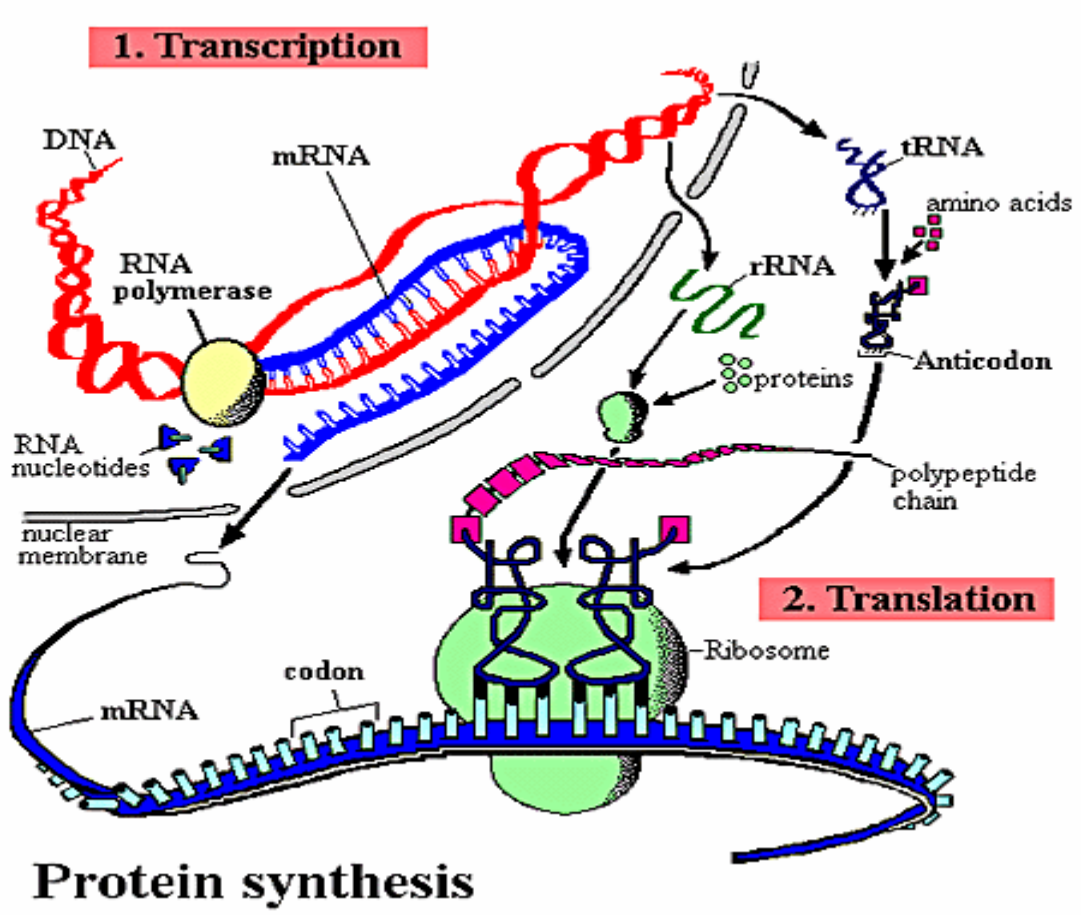


*Figure from *[16]*

2.3) PROTEIN STRUCTURE

Proteins have four orders to their structure: primary, secondary, tertiary, and quaternary. Primary order protein structure is the one-dimensional order of covalently bonded amino acids. This order is a direct result of how the ribosome and the order of codons within the mRNA cooperatively construct the amino acid sequence. There are 20 natural amino acids that can be used in the building of the first order structure of a protein. Immediately after production by the ribosome, proteins are allowed to fold into higher order

conformations. Secondary order protein structure involves the formation of localized structures, most commonly α-helices and β-pleated sheets. These structures are the result of hydrogen bonding primarily between the carboxyl and amino groups of the peptide backbone. Tertiary order protein structure is the folding of large sections of the peptide sequence. The interactions that cause this are between the functional groups of the individual amino acids. Covalent disulfide bonds are sometimes created between cysteine groups, greatly stabilizing the protein structure. In water-soluble proteins, hydrophobic residues tend to group into the middle of a protein, leaving more hydrophilic residues exposed on the protein's exterior. In hydrophobic proteins, the hydrophilic residues group toward the protein's interior leaving the hydrophobic residues exposed. Quaternary protein structure involves the grouping together of two or more polypeptides via noncovalent interaction to form an oligomer, such as the hemoglobin tetramer. Not all proteins have quaternary structure. For a protein to be able to perform its cellular function, all orders of its structure must be precise. Factors such as temperature, pH, and the surrounding solvent can alter the higher order structure of a protein, thereby determining its biological activity or inactivity. In 2D-LC-MS/MS, it is important to conserve only the first order protein structure during isolation, extraction, and separation, because MS cannot determine higher order structure. However, the way a protein's structure is changed must be consistent from run to run in order to ensure consistent digestion with enzymes and consistent retention in LC.

2.4) BIOLOGICAL ROLES OF PROTEINS

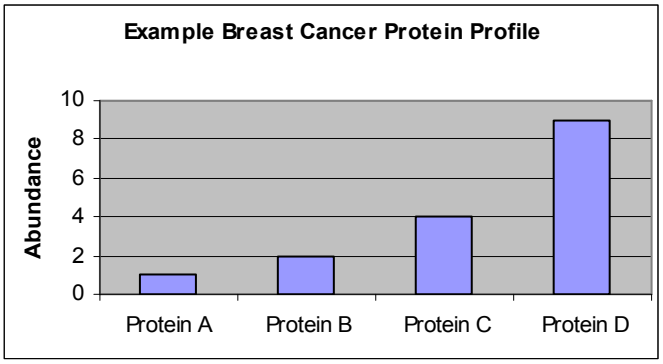
Proteins are a useful class of molecules to study, because they carry out most of the functions within a cell. Proteins are involved in many roles, including O_2 transport, nuclear arrangement, transcription, DNA replication, translation, metabolic pathways, carbohydrate synthesis, and as enzymes *[15]*. It is because of this that differential expression of proteins can sometimes be correlated with the existence of certain diseases such as cancer.

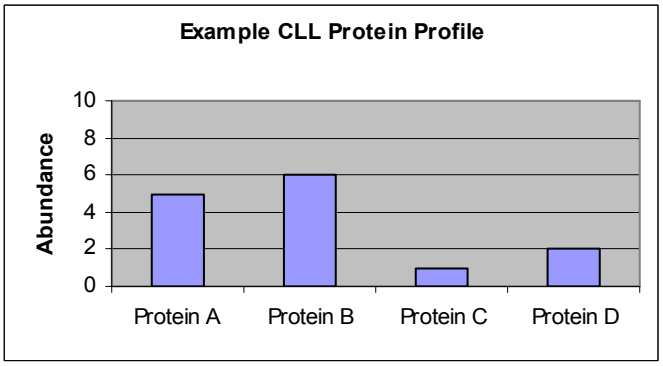
2.5) PROTEIN EXPRESSION DEPENDENCY ON DISEASE

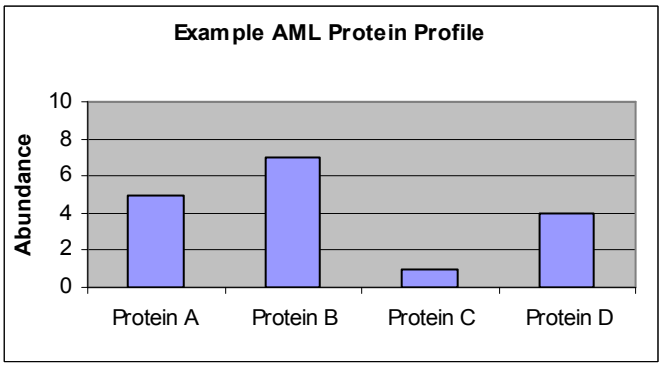
The protein expression within a cell is highly dependent on the specialization of the cell and therefore the cell's location within the body. In addition to protein expression differences due to cell specialization, the protein expression changes from normal to disease characteristic in certain disease states. This change is in part due to the change in cell specialization due to the manifestation of a disease, because the cell's surrounding environment and influences change in many disease states. It has been shown that various types of cancer are accompanied by changes in protein expression and/or changes in modification of certain proteins. A prime example of this is the change in histone acetylation in Chronic Lymphocytic Leukemia and Acute Meyloid Leukemia *[21]*. The offspring of each cell may also be influenced during cell division to conform to the disease state. An example of this is a tumor. Therefore, the protein expression from an entire area of cells in a patient with a disease may change from normal to disease-characteristic expression. This fact enables the use of analytical methods, namely 2D-LC-MS/MS, for the determination of dependencies of protein expression on various diseases and establishment of protein profiles. The following figure shows highly

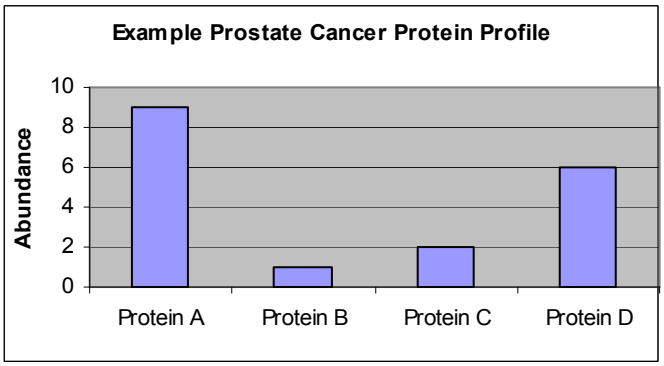
simplified examples of how protein profiles might vary among patients with various cancers.

FIGURE 2.5.1: Example Relationships Between Disease and Protein Expression









3) PROTEIN ISOLATION

Before 2D-LC-MS/MS can be used to characterize and identify proteins, proteins must be isolated from the patient. It is crucial that the methods used to isolate proteins from the patient conserve the proportionality and relative abundances of proteins that exist in the tissue or serum. Many factors must be considered in cell lyses, protein extraction, protein digestion, and all other sample preparation steps. Since 2D-LC-MS/MS analyses are generally concerned only with the masses and sequences of polypeptides and proteins, the secondary and tertiary structures of these molecules need not be maintained in order to conduct faithful analyses. Primary structure must be maintained, as this is what is being analyzed. However, to maintain constancy of retention factors in LC, care must be taken to ensure that, for each polypeptide, there are no structural changes from run to run. Additionally, the quaternary structure of proteins must be considered during analysis. Because in-vivo protein higher-order structure need not be maintained, moderately harsh cell lysis and extraction techniques can be employed. Since mammalian cells have no cell walls, their weak outer membranes can easily be disrupted *[4]*. The lysis of mammalian cells and the techniques used for lysis are not a major limiting factor for most protein extraction procedures. Therefore, a wide range of lysis techniques can be used to suit the needs of the extraction.

3.1) CELL LYSIS

After cells are purified from a patient sample, they must be lysed before proteins can be extracted. Common lysis techniques used for eukaryotic cells include osmotic shock, the use of detergent, enzymatic digestion, various types of homogenization, and grinding

with beads *[4]*. These techniques each have inherent drawbacks and advantages. The technique of choice can be chosen depending on the type of cell to be analyzed and so that the lysis technique is compatible with the subsequent protein extraction technique. Osmotic shock, which breaks the exterior cell membrane by creating a level of osmosis that the membrane cannot tolerate, is unlikely to destroy the primary structure of any proteins and is unlikely lower any protein yield. Grinding with beads and homogenization are relatively harsh, but are unlikely to negatively affect the yield of any proteins. Using detergent disrupts the cell membrane by forming micelles with the membrane and removing sections of it. Detergent methods are very likely to lower the yield of hydrophobic proteins, because many hydrophobic proteins, both within and attached to the cell membrane, are removed along with the micelles. Enzymatic digestion as a lysis technique should be avoided, because the enzyme used may digest many of the proteins of interest, altering their primary structure. If enzymatic digestion is to be used at all, it should be saved for downstream, where digestive enzymes can be used to systematically break down the proteins of interest in order to facilitate their primary structural analysis. Therefore the methods of choice for animal cell lysis are osmotic shock, homogenization, and glass bead grinding, depending on the type of sample being analyzed—fibrous tissue samples may require homogenization for destruction of extracellular connections.

3.2) PROTEIN EXTRACTION

Different parts of an animal cell have different types of proteins. Membranes within the cell contain mostly hydrophobic proteins. The nucleus of a cell contains many basic

proteins *[9,10,22]*. The cytoplasm contains many hydrophilic proteins that can be either basic or acidic. The properties of a protein vary depending on the location and function within the cell. After a cell is lysed, the next step is extract proteins from the cell lysate. Major limiting factors for protein extraction procedures arise when multiple types of protein (i.e. - differences in hydrophobicity, pKa, or buffer solubility) are being extracted at one time. Usually, an extraction procedure will focus on a select group of similar proteins, rather than trying to isolate all cellular proteins at once. For example, to isolate the very hydrophobic membrane proteins, various micelle-forming detergents are used. These micelles solubilize the hydrophobic proteins but do not solubilize hydrophilic proteins. To isolate nuclear proteins such as histones, various buffers are used to isolate cell nuclei, and an acidic reagent is then used to extract the selected basic proteins with minimal solubilization of DNA. Beyond the differences in hydrophobicity, the pKa's of proteins are a major factor in determining their solubility. At a high pH, low functional pKa proteins will dissolve, and at a low pH, high functional pKa proteins will dissolve (the functional pKA is defined as the overall pKa of the protein). Selectively dissolving proteins in an acid or base works efficiently for proteins containing many basic residues, lysine (K) and arginine (R), or many acidic residues, glutamic acid (E) and aspartic acid (D), because these groups have very high and very low pKa's, respectively. Basic proteins contain a high amount of K and R amino acids, and acidic proteins contain a high amount of D and E amino acids. The following figure shows the deprotonation and hence changes in solubility for several groups of proteins. Note that protein solubility is maximized when the protein molecule is most highly charged (when peptide residues

within the protein are charged) and that the peptide reactions shown can occur in multiple sites within a single protein.

FIGURE 3.2.1: Equilibria of Peptides and pH

Basic Protein (high pKa):

$$peptideH^+ \rightleftharpoons peptide + H^+$$

low pH----------high pH

Intermediate Protein (intermediate pKa):

$$peptideH^+ \rightleftharpoons peptide + H^+ \quad \textbf{OR} \quad peptideH \rightleftharpoons peptide^- + H^+$$

low-intermediate pH (~6) intermediate-high pH (~8)

Acidic Protein (low pKa):

$$peptideH \rightleftharpoons peptide^- + H^+$$

low pH----------high pH

Extraction buffers' salt concentrations also determine a polypeptide's solubility due to changes in protein hydration with changes in the buffer's ion concentration. Usually, as the ion concentration of a buffer is increased, protein solubility is decreased. The reason protein solubility decreases with increasing salt concentration is that the salt ions displace water molecules around the protein molecules and thereby compete with hydration of the protein. This phenomenon is shown in terms of protein concentration vs. salt concentration in the following figure.

FIGURE 3.2.2: Protein Solubility vs. Salt Concentration*

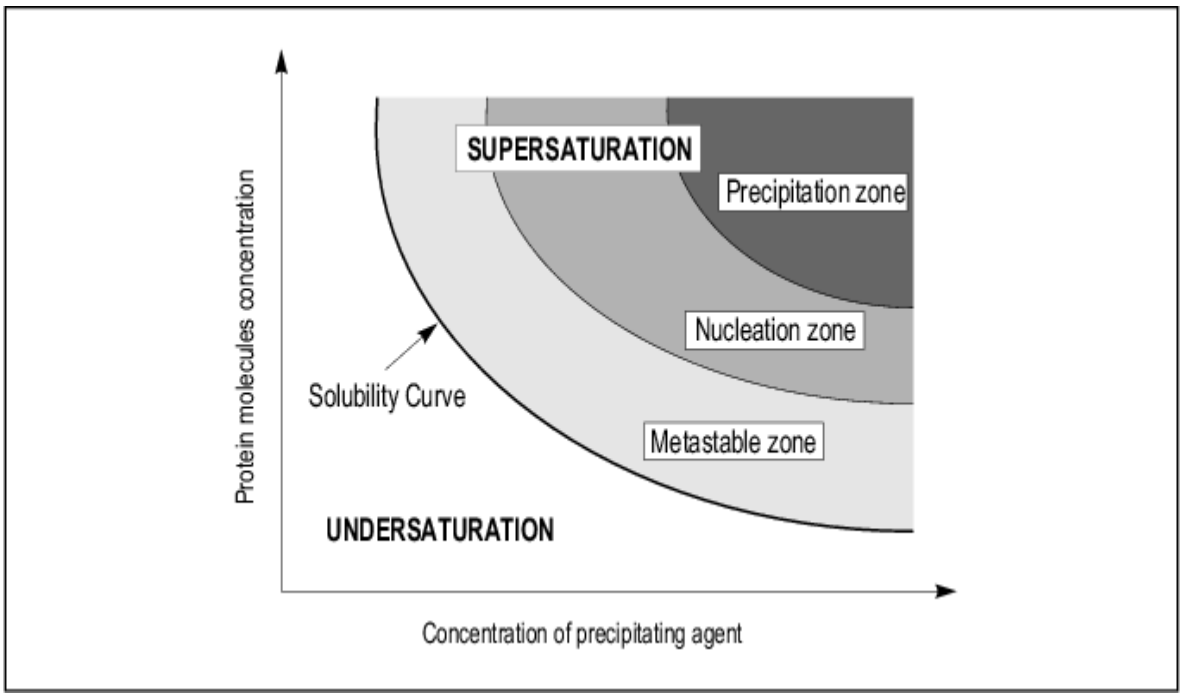


*Figure from *[11]*

To extract every protein from a cell, it is necessary to sweep a polarity gradient, a buffer concentration gradient, and a pH gradient. Exposing all proteins within a cell lysate to all possible conditions (within appropriate ranges) will ensure that each protein has a chance to dissolve and thus be extracted. The range of pH values to be tried should be roughly 2-12, avoiding extremes in order to minimize unselective solubilization; many proteins are solubilized at extremes of pH *[23]*. The polarity gradient should consist of 100% water on one end and 100% of a much less polar solvent (i.e. - ethanol, propanol, or hexanol) on the other end. The buffer concentration gradient should start with 100% water and end with roughly 1 M ammonium acetate. It is important that all solvents and buffers be removable in order to be compatible with 2D-LC-MS/MS. Dialysis

techniques, in which there is a pore size larger than the salt ions but smaller than the proteins, can later be used to remove salt from the proteins.

3.3) PROTEIN DIGESTION AND SAMPLE PREPARATION

This section investigates the details of protein digestion and discusses whether digestion need be used. Digestion has advantages and disadvantages with regard to 2D-LC-MS/MS analysis. On one hand it fragments proteins extensively, thus assisting in a top-down MS analysis of the resultant peptide fragments so that the exact location of modifications with a protein can be pinpointed. On the other hand, the digestion can lead to relatively poor separation in cation-exchange chromatography and can be unpredictable. In many cases protein digestion facilitates the analysis of the primary structure of large proteins. In many cases an enzyme that always cleaves a protein at a certain amino acid, or type of amino acids, is used. Trypsin is one such enzyme that cleaves proteins at positively charged residues, namely lysine and arginine. Trypsin works well for basic proteins, which are full of these positively charged residues. Due to the secondary, tertiary, and quaternary structures of many proteins, not all expected sites are cleaved by enzymes such as trypsin. At very low concentrations of trypsin, the concentration is vaguely proportional to the amount of cleaved sites.

For a given enzyme there is a focused type of protein that will be fragmented more than most others. In the case of trypsin, which cleaves at positively charged sites, basic proteins are fragmented most. Since trypsin is not completely efficient with respected to the number of expected sites cleaved, the charge distribution of the resulting fragments

from a highly basic protein might be something like: …-1, 0, +1, +2, +3,+4… at pH~7 with the distribution being focused primarily at +1 and tapering off for charges further from +1. The enzyme endoproteinase GluC cleaves proteins at negatively charged residues, namely glutamic acid and aspartic acid *[5,6]*. For endoproteinase GluC, the proteins that will be fragmented most are the acidic proteins, which contain more negatively charged residues than other proteins. Since GluC also has limited efficiency, the charge distribution of the fragments from a very acidic protein digested by GluC might look something like: …-4, -3, -2, -1, 0, +1… at pH~7 with the distribution being focused primarily at –1. For proteins that are neither very basic nor very acidic, the charges of the resulting fragments from a digestion with trypsin or GluC are more likely to be neutral as the positive and negative residues' charges of a protein of neutral pKa are likely to cancel each other.

Since the many proteins in a cell vary widely in pKa, the results of a trypsin digestion on a solution of all proteins from a given cell will vary greatly from protein to protein. If all of the proteins are first digested and next have their pH lowered to 2, the resultant fragments will likely fall, for example, within the charge range 0 to +5, depending on the cleavage efficiency. Note that the minimum charge here is 0, because it is assumed that if the pH is reduced to 2 all negatively charged residues will be protonated. If none of the proteins are digested, the charge distribution of the proteins at pH~2 will likely range from 0 to +N, where N is the maximum combined number of arginine and lysine residues in a single protein molecule. Since N is an order of magnitude or more greater than the highest charge on any one of the fragments resulting from digestion, the charge

distribution for undigested proteins is much wider than that for digested proteins. Since the first dimension of 2D-LC-MS/MS uses cation exchange chromatography, better separation will be achieved if a wider charge distribution of proteins is present. The following figures show hypothetical charge distributions for digested and undigested global protein samples at pH=2.

FIGURE 3.3.1: Model Charge Distribution for Digested Global Protein Sample

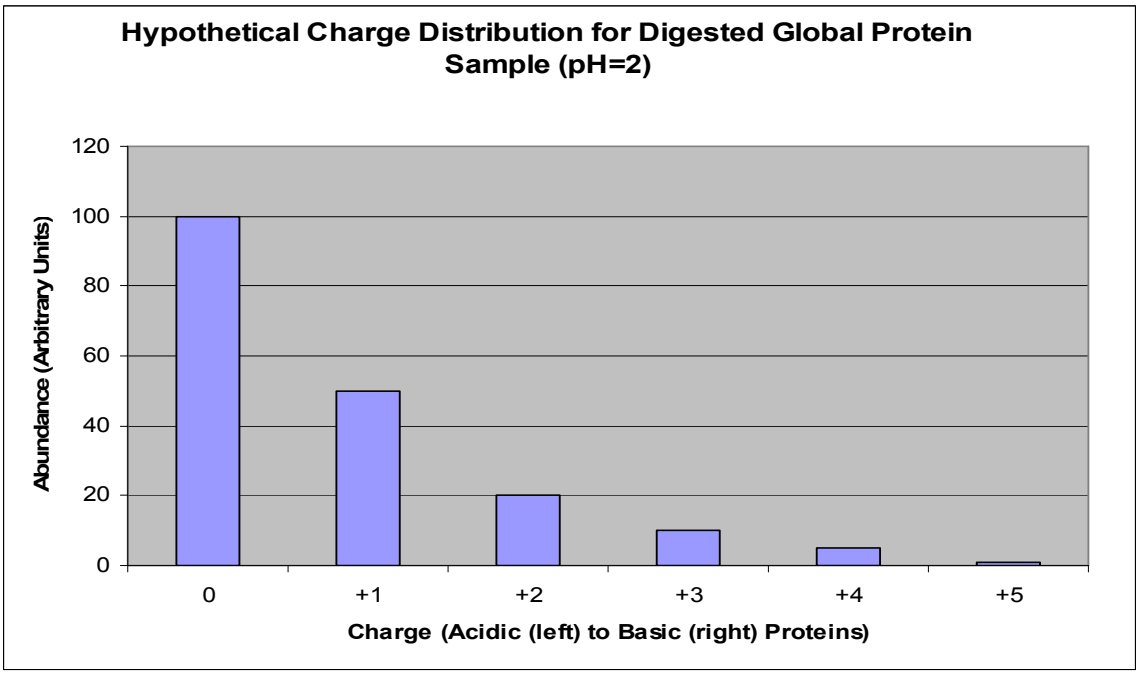
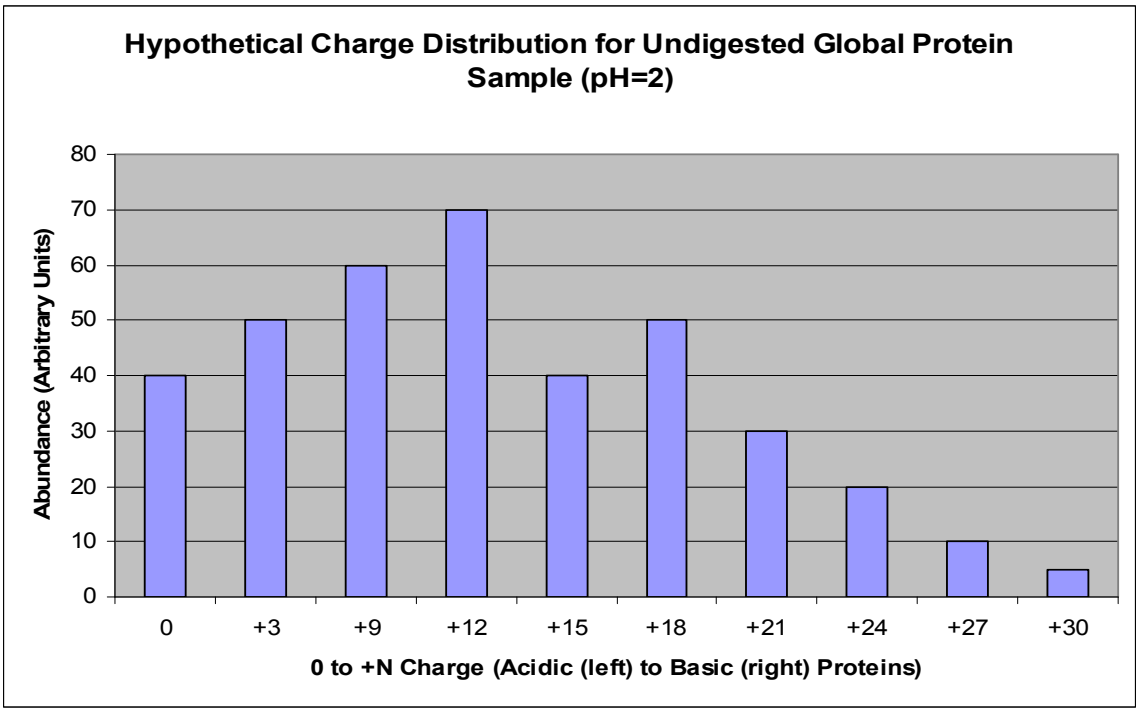


FIGURE 3.3.2: Model Charge Distribution for Undigested Global Protein Sample



Though protein digestion is useful in many techniques, there are several reasons why digestion need not be used in the process of establishing protein profiles. As mentioned previously, cation exchange chromatography has a wider charge distribution from a global protein sample to separate if the proteins are left undigested. Digestion of proteins is unpredictable, as many expected cleavage sites may remain uncleaved. Also, the ability of protein profiles to identify diseases is likely to depend less upon the exact location within proteins of modifications and is likely to depend more on the abundance of modifications and proteins in general. Therefore, establishment of reliable protein profiles via 2D-LC-MS/MS can be achieved best without digesting proteins.

4) INSTRUMENTATION

4.1) LIQUID CHROMATOGRAPHY

Chromatography in general involves separating different species of analyte by exploiting various properties of the analyte. All chromatographic separations involve an analyte, a mobile phase, and a stationary phase. The analyte is the sample to be separated. The mobile phase is the solution the analyte is mixed with and transported by. The stationary phase is usually composed of a solid medium with which the analyte can interact. The mobile phase is usually varied so that at first the analyte is attracted more to the stationary phase than to the mobile phase solution. As the run proceeds, the mobile phase becomes more accommodating to the analyte. The disparity in properties among the analyte is necessary for separation to occur. Properties typically exploited in chromatography include hydrophobicity, charge, affinity, size, and others. The retention time of an analyte describes the comparative strength of the analyte-stationary phase interaction relative to the analyte-mobile phase interaction.

Some key equations and terms involved in chromatography are:

1) Retention time = time after injection for analyte to reach the detector
2) Distribution Constant (K) = (concentration of analyte in stationary phase)/ (concentration in mobile phase)
3) Selectivity Factor (α) = (Distribution constant for standard analyte)/ (dist. const. for other analyte). This term describes how easily two analytes can be separated.

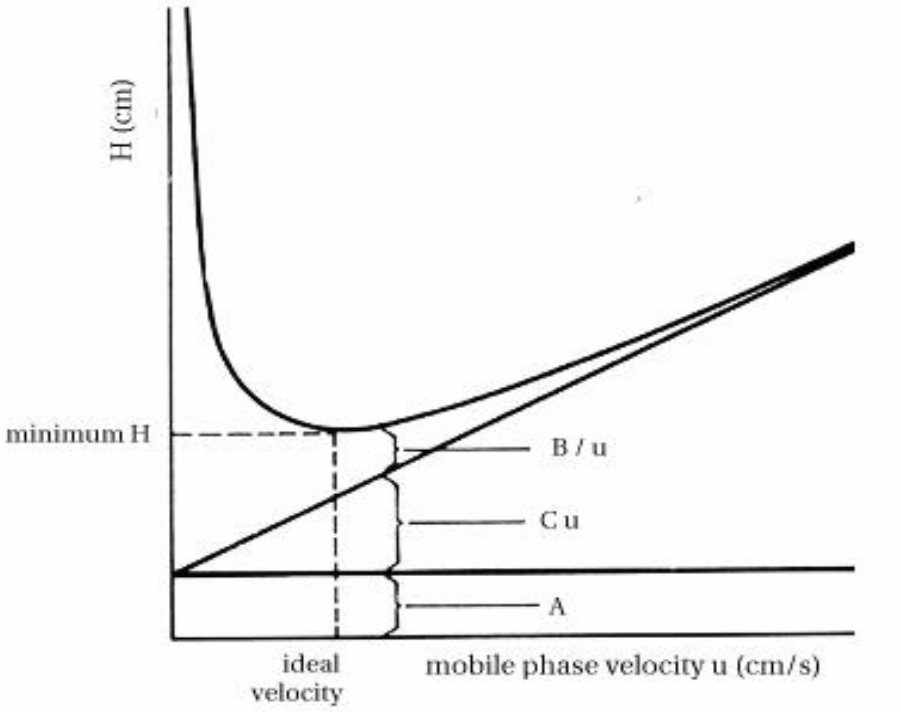
4) Number of theoretical plates (N) = (length of column)/ (theoretical plate height). The number of theoretical plates is a measure of the efficiency of separation and is historically derived from the number of cells in a distillery column.

5) Retention factor (k) = K*(Volume of analyte in stationary phase)/ (volume in mobile phase)

6) Resolution (R) = $\sqrt{(N)} / 4*(\alpha -1)/ \alpha *k/(1+k)$. Resolution is defined as the separation between two peaks. Two adjacent peaks that barely meet have a resolution of R=1.

7) Van Deemter Equation: $H = A + B / u + C * u$; H is the plate height, u = the linear velocity of the mobile phase, A = multiplicity of flow paths, B = longitudinal diffusion, and C = mass transfer between phases. This equation links some typical chromatography variables to chromatic inefficiency. The following figure displays the effects of each term to the plate height, which is inversely proportional to column inefficiency

FIGURE 4.1.1: Van Deemter Plot – Plate Height vs. Solvent Flow Rate*



*Graph from *[19]*

As can be seen in Figure 4.1.1, four terms contribute to the plate height. 'A' describes the mobile phase's and analyte's ability to wind through multiple paths as they flow through a packing material. 'A' is constant for a given column and decreases as columns with smaller inside diameters are used. Because of this fact, column diameters as low as 10 μm are used. 'B/u' describes the diffusion along the column length of the analyte from the analyte's elution zone into more dilute areas in front of and behind the zone. The effect of 'B/u' decreases as the flow rate of the mobile phase increases and as column length is decreased. 'C*u' accounts for the fact that analyte mass can transfer between the stationary phases and the mobile phase occurs, and this process's equilibrium speed is usually slower than the flow rate will allow for. Therefore 'C*u'

accounts for the level of disequilibrium of the mass transfer. 'C*u' has more of an effect at higher flow rates, because the equilibrium is more offset *[14]*.

4.2) MASS SPECTROMETRY

Several types of mass spectrometry (MS) exist. Most applicable to protein characterization among these are ion-trap MS, Fourier transform ion cyclotron resonance (FT-ICR) MS, and time-of-flight (TOF) MS. TOF-MS is typically used in experiments where the sample permits being immediately desorbed from a medium and ionized, as in matrix assisted laser desorption ionization (MALDI). MALDI-TOF MS is very time consuming, because the protein sample needs to be separated prior to MALDI-TOF MS using a separate technique. FT-ICR MS is useful in that it has detection limits down to attomolar concentrations of proteins, has very high sensitivity, and can identify up to 20,000 components in a single mixture without prior separation *[12]*. FT-ICR MS is powerful enough to resolve isotopic distributions of analyte. However, this quality of resolution is not necessary for establishing protein profiles, because differences between modified and unmodified proteins and differences between non-similar proteins can be resolved well enough with less resolution. Ion trap MS is the MS technique most commonly used in 2D-LC-MS/MS due to its high stability and sufficient mass resolution (1 Da), limits of detection, and sensitivity *[14]*.

After chromatographic separation, protein ions are introduced to the mass spectrometer via electrospray ionization (ESI). ESI works by having a high voltage (1-2 kV) potential applied to the protein-occupied solution at a capillary opening. The capillary tubing's

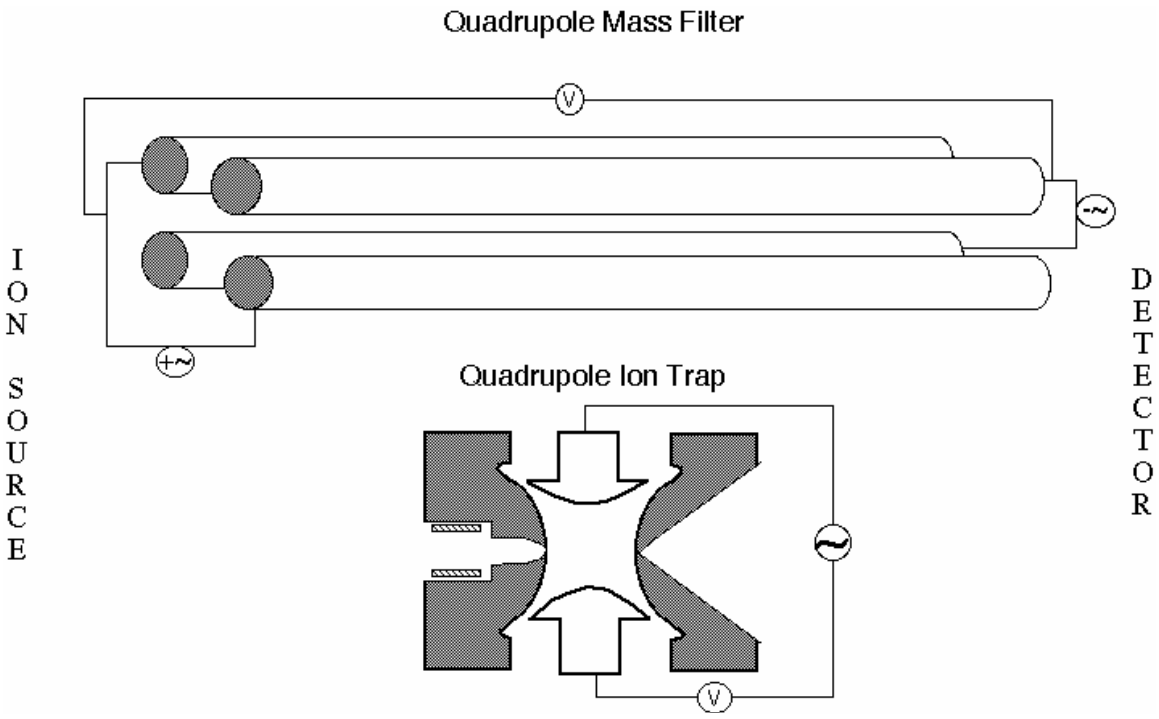
inside diameter should be decreased to such a small size that the capillary opening sufficiently resembles a point charge. Decreasing the diameter of the capillary opening increases the efficiency of ionization of the solution, because there is more charge per unit volume. ESI is a necessary interface between chromatography and MS, because it converts proteins into highly charged forms and desolvates the proteins. It is necessary that the proteins have a high charge, because MS is selective of mass/charge (m/z) ratios. If none of the proteins were charged, there would be no m/z ratio, and more importantly, there would be no selectivity of mass analysis between different proteins. It is necessary that the proteins be desolvated, because having solvent molecules variably attached to the proteins creates much more complex mass spectra and becomes very confusing to deconvolute.

Just as chromatography is selective of certain molecular properties, MS is selective of the m/z ratio of the ions introduced into the MS system. An ion trap MS uses two ring electrodes, which are charged with an alternating radio frequency (RF), to excite the analyte ions. To either side, between the ring electrodes, are grounded electrodes, usually referred to as end-caps. An RF range is scanned so that at lower RF, lighter ions oscillate with a stable orbit, and at higher RF, heavier ions oscillate with a stable orbit *[14]*. As ions' orbits become unstable, they collide with one of the end-caps, causing an electrical discharge. The current from this discharge is detected, multiplied using an electron multiplier transducer, and measured using a computer to form a signal for each ion *[14]*. Comparison of each ion signal with the RF value at which the ion collision occurred provides the user with an abundance vs. m/z spectrum. To obtain MS/MS data for a

given ion, one can freeze the RF scan in the quadrupole analyzer at a particular RF at which the parent ion of choice has a stable oscillation. An inert gas such as helium is introduced, and collisions between the parent ion and the gas molecules occur, causing the ion to fragment into smaller daughter ions. These daughter ions can then be measured by sweeping the ring electrodes through an RF range and collecting m/z data in the same manner as for the parent ion. MS/MS data can be collected for a given ion within a small number scans. The ease with which one can fragment a certain m/z protein and collect daughter ion mass spectra with an ion trap MS makes ion trap MS the method of choice for establishing protein profiles and related daughter ion data.

Most ion trap MS instruments contain a quadrupole mass analyzer that proceeds the ion trap analyzer. The quadrupole mass analyzer works in a method similar to that of the ion trap. In a quadrupole, there are four electrically conductive rods, each exposed to an alternating RF frequency. Rods opposite from each other are electrically connected. Two of the rods are connected to the positive side of a direct current source, and the other two rods are connected to the negative side of the direct current source. Alternating RF currents are exposed to the two pairs of rods and are 180° out of phase with each other *[14]*. As in ion trap MS, only ions with a certain m/z value remain in a stable orbit at a given moment in time. In most MS/MS experiments, the quadrupole analyzer is used to focus on a parent ion, and the ion trap analyzer is used to analyze the subsequent daughter ions. The following figure shows a quadrupole mass analyzer and an ion trap mass analyzer:

FIGURE 4.2.1: Ion Trap and Quadrupole Mass Analyzer Diagrams*



*Picture from *[20]*

4.3) 2D-LC-MS/MS

2D-LC-MS/MS incorporates two types of chromatographic separation and tandem mass spectrometry in one automated process. The two types of separation are ion exchange and reversed-phase (RP) separation. The type of ion exchange chromatography used depends on the pH of the mobile phase used during separation. For low pH, strong cation exchange (SCX) chromatography should be used, because most proteins are protonated and are thus positively charged (cations) at a low pH. Anion exchange chromatography is used if the mobile phase is held at high pH, because most proteins are deprotonated and are thus negatively charged (anions) at high pH. In SCX, analyte in the form of cations is introduced to a chromatographic column with immobilized anionic functional groups.

The cations have affinity for the immobilized anionic functional groups and are thus retained on the column. As increasing concentrations of salt are introduced to the column, salt cations compete with the analyte cations for binding to the anionic column, and analyte cations are selectively eluted. Analyte cations with lower charge are eluted first. Selectivity of the analyte cation elution is primarily based on differences in charge. Further selectivity of analyte cation elution is based on size and polarity. 2D-LC-MS/MS will be discussed in terms of SCX with mobile phase solvents at low pH. In RP chromatography, the second separation dimension of 2D-LC-MS/MS, the analyte is injected onto the column, and the solvent generally shifts from water (most polar) to an organic solvent (less polar). As the solvent gradient shifts toward less polar, the analyte desorbs from the stationary phase and elutes, more polar analyte eluting first.

Typical mobile phase compositions used in RP separation of proteins or peptides are: Solvent A, water with .1% formic acid; Solvent B, acetonitrile with .1% formic acid. Trifluoroacetic acid (TFA) is used in place of formic acid (also at .1%) in many cases. TFA or formic acid, which lower pH, are added to the mobile phase in chromatography, because their anionic species bind (ion-pair) with the cationic protein species (many proteins are cationic at low pH), thereby neutralizing proteins, markedly decreasing solubility, and increasing column retention. If ion-paring species such as formic acid or TFA were not used, most cationic proteins would be lost. TFA induces better separation through this mechanism than does formic acid. However, formic acid is much more compatible with MS, because it is more easily desolvated during ESI, allowing positive ions to form. The lowering of pH that accompanies the use of an ion-pairing agent

should also lower the solubility of acidic proteins, thereby increasing retention of as many proteins as possible.

Typical mobile phase compositions used in SCX separation of proteins or peptides are: Solvent A, water; Solvent B, 500mM ammonium acetate in water. In addition to having a suitable cation upon dissociation, ammonium acetate is chosen because it is sufficiently volatile that it does not cause significant interference in MS. It is likely that the use of TFA and formic acid have devastating effects in SCX separation, because their ion-pairing action will neutralize many of the positive charges of the proteins, thereby eliminating protein affinity for the anionic SCX column. The facts that ion-pairing agents are necessary in RP chromatography and that ion-pairing agents cannot be used in SCX chromatography, presents a major incompatibility between the two methods.

To overcome this incompatibility between SCX and RP separation, four pump systems can be used where SCX separation occurs without an ion-pairing agent being present, and RP separation occurs with the ion-pairing agent present. This method allows the analysis of peptides resulting from protein digestion. The following figure depicts a typical four pump 2D-LC-MS/MS system.

FIGURE 4.3.1: Four Pump 2D-LC-MS/MS Schematic*

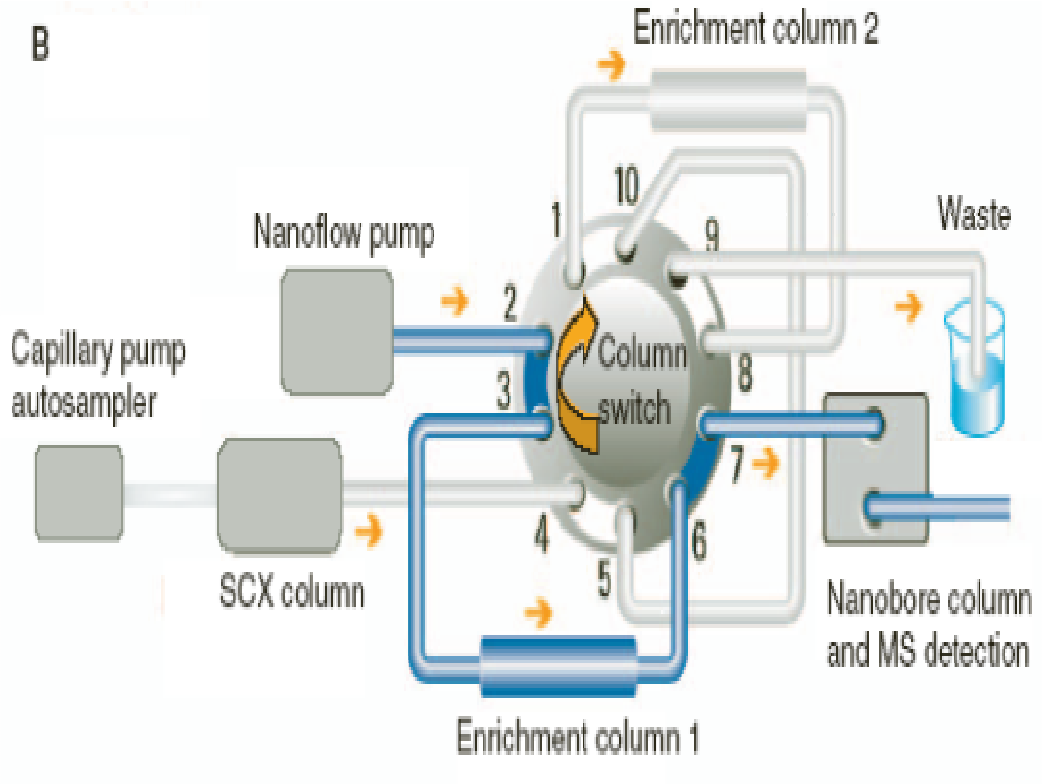
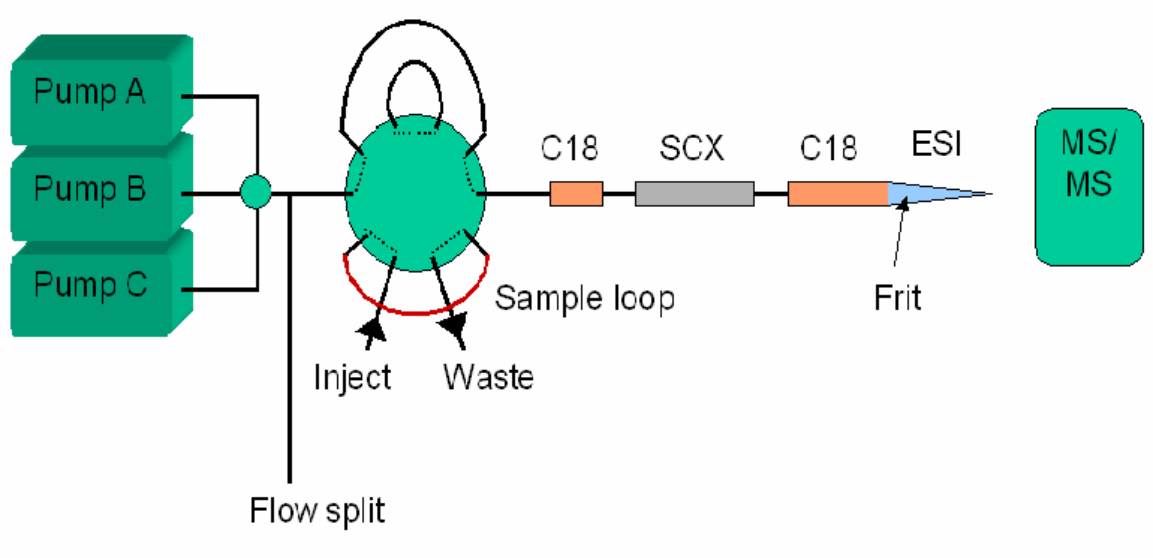


*Figure from *[15]*

Another possible method requires that the proteins to be analyzed are not digested. The idea here is that the charge distribution of proteins will be so wide and contain high enough protein charges that the presence of an ion-pairing agent will not significantly shrink the charge distribution of proteins. This method allows the use of only three pumps, because the addition of a fourth pump with water free of ion-pairing agent is not necessary. This reduces overall costs in owning and operating a 2D-LC-MS/MS system and greatly simplifies the method used. The following figure shows the 2D-LC-MS/MS

system in which: proteins should not be digested prior to analysis, only three pumps are needed, and ion-pairing agents can still be used. The mobile phase compositions in this picture are: Solvent A, water with .1% formic acid; Solvent B, acetonitrile with .1% formic acid; Solvent C, 500 mM ammonium acetate in water.

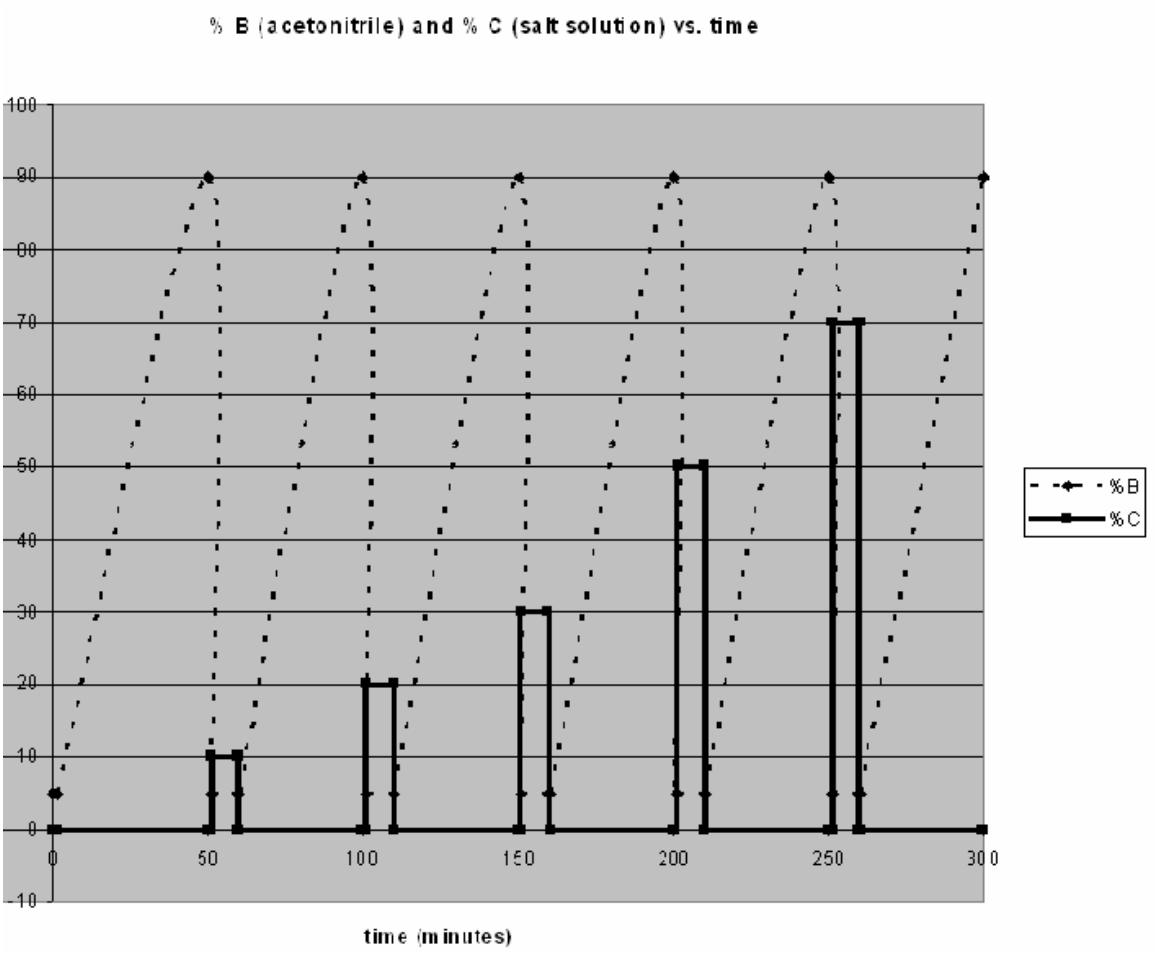
FIGURE 4.3.2: Three Pump Linear Column 2D-LC-MS/MS Schematic



In this system, the analyte is loaded in solution into the sample loop, which ensures that the sample injection volume is the same from run to run. The analyte is then washed in the first C18 column so that all salts from extraction procedures are removed from the sample. An increase in acetonitrile concentration then washes the analyte to the SCX column, where the cationic proteins are retained. At this point, the salt concentration (Solvent C) in this 2D-LC-MS/MS system is increased in a stepwise fashion to elute in steps cationic proteins of increasing charge from the SCX column. After each step in salt concentration increase, a polarity gradient starting with ~95% water and ending with

~90% acetonitrile is intermittently run. This polarity gradient separates proteins on the final C18 column. The following figure shows an example of a 2D-LC-MS/MS gradient.

FIGURE 4.3.3: 2D-LC-MS/MS Solvent Control Diagram



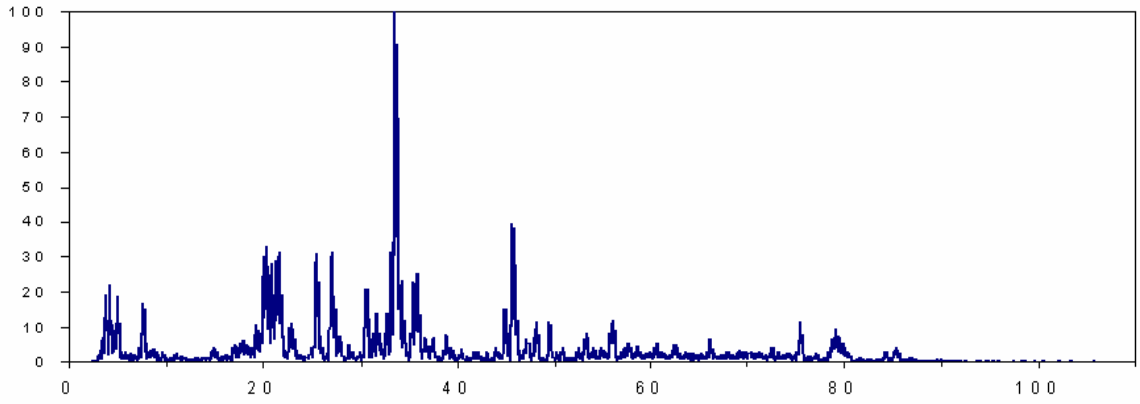
4.4) ADVANTAGES OF 2D-LC-MS/MS OVER LC-MS/MS

2D-LC-MS/MS has some important advantages over one-dimensional liquid chromatography - tandem mass spectrometry (1D-LC-MS/MS or LC-MS/MS) in the analysis of very complex mixtures such as protein mixtures obtained via global protein extraction from cells. Generally speaking, 2D-LC exploits more properties of protein

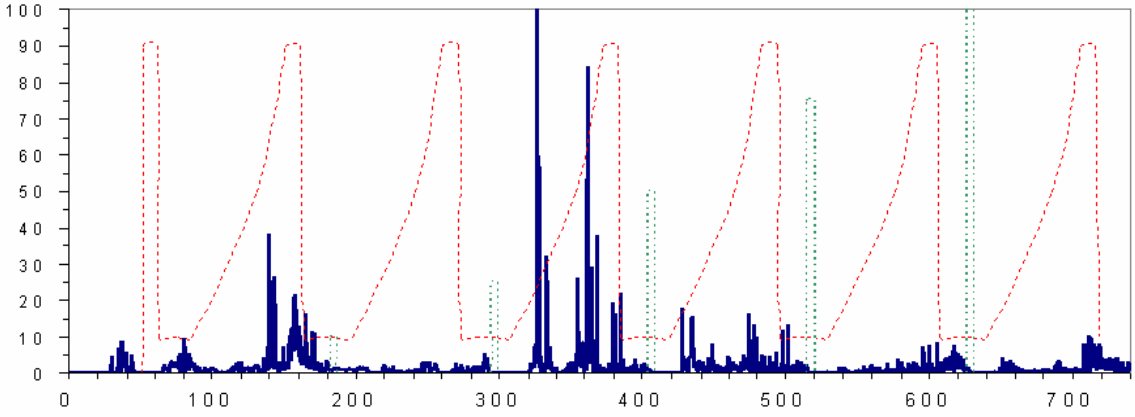
mixtures and peptide mixtures than does 1D-LC. Whereas RP chromatography only utilizes differences in polarity and size of the analyte, 2D-LC exploits these features and differences in the charge among the analyte. The result of having more methods of separation is having a higher amount of separation. A mixture of proteins that is grouped together as one peak in 1D-LC is likely to be separated into multiple peaks in 2D-LC. Perhaps the only disadvantage of 2D-LC is the large amount of time required for a single run. However, when considering that each protein to be separated requires a certain amount of time, or peak width, and that more time is used in 2D-LC, the peak capacity in 2D-LC is on the order of magnitude of 10X higher than in 1D-LC. The advantages of the higher degree of separation and increased peak capacity in 2D-LC far outweigh the disadvantage of longer run time. The results of having a higher degree of separation are that it becomes much easier for MS and MS/MS to resolve proteins from one another and to identify which daughter ion fragments are from which parent ion proteins, respectively. Chromatograms of trypsin digested peptide fragments separated using 1D-LC-MS/MS (figure 4.4.1-a) and 2D-LC-MS/MS (figure 4.4.1-b) are shown below. Figure 4.4.1-b also shows the salt gradient (spikes) and acetonitrile gradient (ramps) in the background. Bear in mind that the 2D-LC-MS/MS chromatogram is stretched over ~700 min, whereas the 1D chromatogram is only ~100 min in duration. Additionally, note that the 2D-LC-MS/MS separation would most likely be better if the original proteins had not been digested, because the charge distribution would be much wider.

FIGURE 4.4.1: Chromatograms for 1D-LC-MS/MS and 2D-LC-MS/MS

a) 1D-LC-MS/MS



b) 2D-LC-MS/MS



5) NON-DISEASE-SPECIFIC VARIABILITY OF SAMPLES

For each disease, there is an optimal place within the human body where the consequential differences in protein expression are most prominent. For example, liver cancer's manifestation can best be seen in the liver, and samples taken from the liver will best show the differences in protein expression due to this disease. Since diseases each have their own ideal sampling locations within the body, there will be a great deal of variability in non-disease-specific protein expression due to the differing specialization of

cells in different areas of the body. When creating protein profiles whose differences correspond only to the effects of each disease, it is necessary to eliminate any differences among profiles that are due to sampling location. One method to solve this problem is to establish blank (disease-free) protein profiles for all areas of the human body that will be tested and establish standard deviation values for all natural proteins at each of these areas. Additionally, blank protein profiles can be created for modifications in the procedure, such as use of different solvents, enzymes (if digestion is preferred), or protein extraction techniques. Once many blank profiles and disease-state profiles are created, statistical validation techniques can be used to determine what differences in protein profiles are due to cell specialization and what differences are due to diseases. When protein profiles from patients are compared with database protein profiles, match results will be most statistically credible for the patient's profile whose procedure was most similar to the procedure most commonly used to obtain database protein profiles. For the patient whose protein profile is obtained using non-standard techniques, the results will be less credible, because there will be fewer database protein profiles whose methods were the same. At first glance this seems like a major drawback, but as more protein profiles obtained by various methods are added to the database, more credibility will be associated with each method.

6) DATABASE AND SEARCH TOOLS

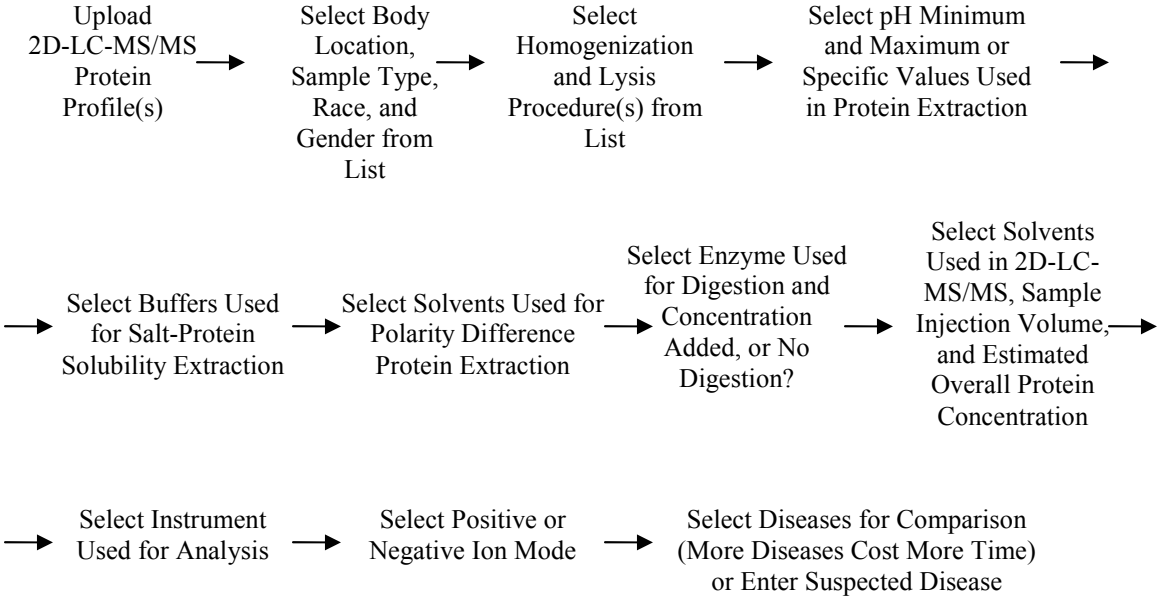
Due to the immense amount of variability in non-disease-specific protein expression and in protein isolation techniques, it is critical that the database and search tools used to handle protein profiles be sensitive to and aware of these experimental differences.

Sources of non-disease-specific variability from experiment to experiment in protein profiles include: different lysis procedures, different sources of sample (i.e. – tissue, serum), different instrumentation, different solvents in LC, different extraction procedures, whether the protein sample was digested, biological differences between subjects such as race or gender, and many other factors. To account for the many differences in methods used, as much information as possible must be included for each protein profile on the database. It is thus necessary to have a program with input values for all of the different parameters that pertain to each experiment. Having a computational method for determining the relevancy of differences of experimental parameters among experiments will provide one with many bases for determining statistical relevancy between profiles. In such a computational method, ranges of numerical values can be assigned to the ranges possible for given parameter. For each parameter there is a group of certain ranges of numbers. The user's input through the user interface will determine what value among the range of possible values will be used for the parameter in question.

As a specific example, consider a case when the user uploads a protein profile and selects "adult red blood cells" within the "Select Body Location, Sample Type, Ethnicity, and Gender from List" option available on the program. For this selection, a numerical sample type value of 3 points is assigned to the protein profile. The range of sample type values extends from 0 to 10 points, where the values indicate, with roughly linear estimation, the amount of distinct proteins available in the sample. Since adult red blood cells have relatively few distinct proteins, a low value of 3 is given to the profile. A

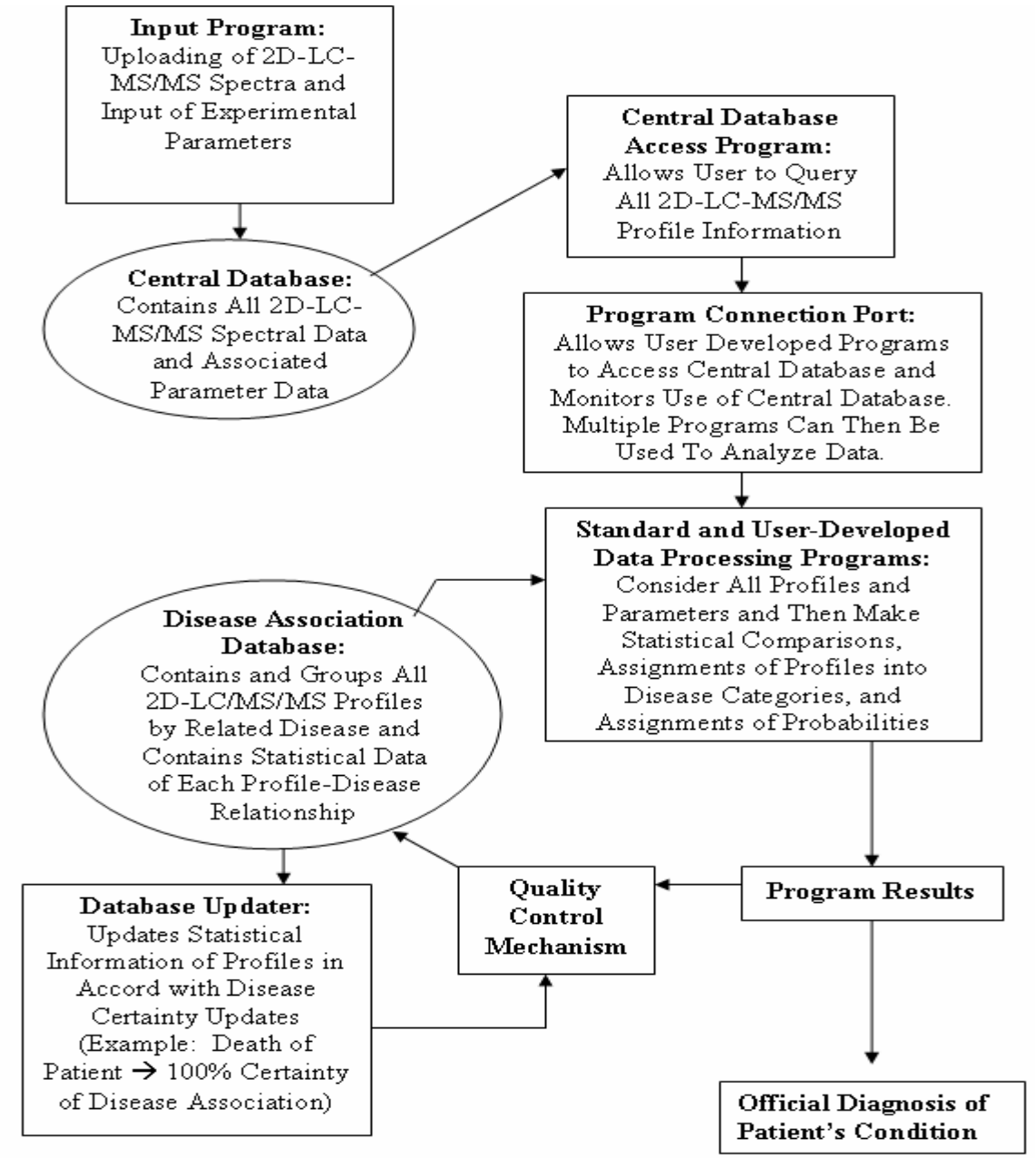
protein profile obtained from skin cells or muscle cells might receive a higher value, such as 7 out of 10 possible points. A profile obtained from glandular cells might receive a value of 10 out of 10 points due to optimal distinct protein content. As the value increases, a higher amount of distinct proteins is expected to be present and a higher signal to noise ratio is thus expected. Therefore the protein profile associated with a higher value can be given a higher confidence level in terms of accurate representation of the disease state in question. In addition to a rating of the amount of distinct proteins, many other factors can be associated with the sample type and can be given respective numerical values and possible ranges. This type of number/range assignment can be given to each parameter input in the program. The following figure shows a rudimentary program for the input of sample and experiment associated data and the uploading of experimentally obtained protein profiles:

FIGURE 6.1: Example Program for Profile Uploading and Parameter Input



The database used to hold the protein profiles will grow as more protein profiles are added. Therefore, for any given method used, for any part of the body sampled, and for any disease, the statistical basis for any decision will increase as time progresses. In order for this to happen, each protein profile compared with the database must eventually be added to the database along with an official determination of the disease that corresponds to each profile. In some cases, the addition of the profile to the database may take awhile, because it is necessary to first determine whether the patient does or does not in fact have the disease. Alternatively, each database protein profile could have a relative certainty value attached to it that describes the amount of certainty that the profile is related to a certain disease. Each time a patient's profile is compared with the database, the amount of certainty of each database profile will determine the amount of association between the profiles for a given disease. The relative certainty value for a profile could later be updated if it was later determined that the patient definitely does or does not have the disease. The following diagram demonstrates one possible method for data flow in protein profiling.

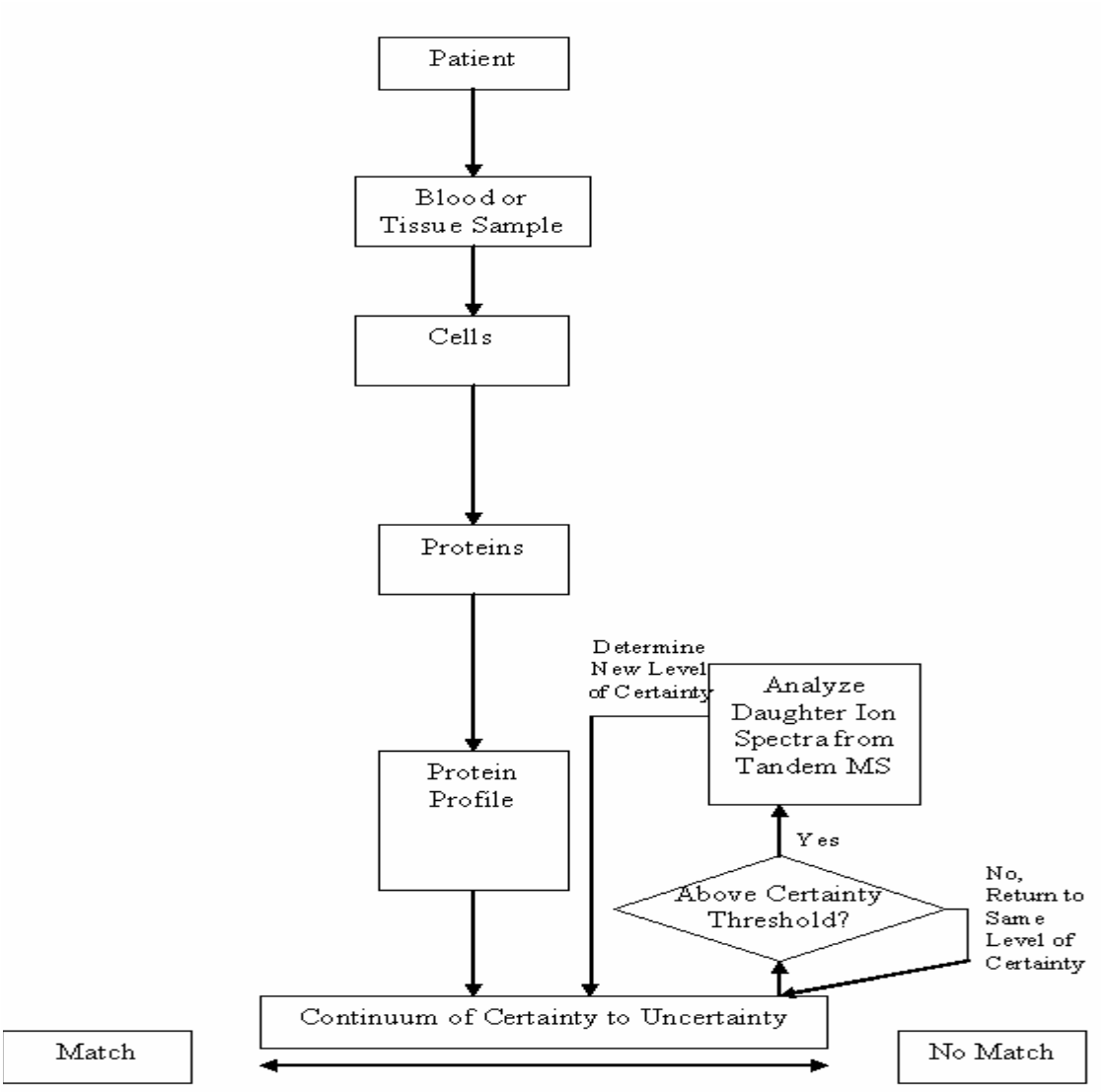
FIGURE 6.2: Proposed Schematic for Database Interface and Data Analysis



To increase the certainty of a profile and create more routes for association between similar profiles, the daughter ion spectra of the original profiles can be used for more in depth analysis. In this scenario, the MS (parent ion) data are used to create the protein profiles. When a match is made for a profile to another group of profiles within a predetermined threshold of certainty, the MS/MS (daughter ion) data from each profile

can then be compared to determine a new level of certainty of the match. Through this method the errors associated with protein profiling can be minimized to a great extent, depending on the amount of MS/MS data available. The following figure illustrates the process by which profiles are obtained and the process by which MS/MS data can be used to increase confidence levels of a given match:

FIGURE 6.3: Flowchart of Information from Patient to Disease Association



7) CONCLUSIONS

In order to easily compare the underlying traits of one disease to those of another, it is necessary to have a standardized method of obtaining the diseases' information. Otherwise, comparison of multiple research groups' data would either be inconclusive or be very difficult to conduct due to the massive incompatibilities of one method to another. Additionally, a single person can understand only a limited amount of information, and fully understanding the results obtained via many different methods is very time and energy consuming. Using bioinformatics is a solution that many research groups are turning to, because letting computers do as much work and data analysis as possible allows scientists to continue leading study and resources into uncharted areas of research. This is a trend that has been observed in the computer industry for the last 40 years or more. As computational widgets achieve higher and higher ability, more and more computational work that would otherwise be done by people is done by computers.

In terms of developing a standardized method for analyzing protein profiles from various diseases, 2D-LC-MS/MS provides many characteristics necessary for such a purpose. 2D-LC-MS/MS is very fast in comparison with traditional techniques such as gel electrophoresis and immunoassay techniques. 2D-LC-MS/MS has the potential to measure a very large amount of proteins from a single sample. 2D-LC-MS/MS provides a large amount of information relative to the information encoded within a given sample. This is possible due to MS/MS analysis of daughter ions, the ability to measure many different proteins, and the ability to use information about retention in 2D-LC in coordination with MS data to statistically validate findings.

The growth and proliferation of the protein profile database and associated programs will likely be characterized by: enormous compilation of protein profiles for each disease, consideration of more and more experimental and sample parameters, addition of new computational values and ranges associated with each parameter, extra MS/MS data, and increased match-certainty determination capabilities. It would not be surprising if such a database grew to contain many terabytes of useful information as research groups and medical institutions all start to contribute protein profiles and profile comparison ideas. It is necessary that the database and associated programs be fluidic, allowing acceptable user-induced changes to be made. Therefore, this database and associated programs would serve not only as a key for identification of certain diseases like cancer but also as a method to uptake and exercise new paradigms and methods in protein analysis. As more and more protein profiles are analyzed, it is likely that many more proteins will be discovered, thereby leading to further coverage of the human proteome.

8) SOME PERSONAL PROTEOMIC RESEARCH CONTRIBUTIONS

8.1) 2D-LC

During the development of the 2D-LC-MS/MS system, multiple considerations needed to be taken, and many problems were overcome. Considerations included what type of column packing materials to use, what plumbing and system control configurations to use, what solvents should be used, compatibility between SCX and reversed phase chromatography, and many others. Some of the problems that arose included lack of analyte retention in the SCX column, salt removal, poor initial separation, and difficulty

obtaining reproducibility. The problem with lack of analyte retention was solved by the addition of a C-18 trap column prior to SCX separation. The trap column removed salt from the sample prior to SCX introduction, and the lack of salt in the sample thereby allowed the analyte to be retained by the SCX column. It was therefore found that salt must be removed prior to SCX introduction, because the presence of salt interferes with the retention of the analyte in the column. Difficulties in obtaining reproducibility were partially overcome by the simplification of the 2D-LC/MS setup. Finding an optimal solvent gradient method increased the degree of separation of peptides in 2D-LC. Many issues were confronted and overcome in terms of making the 2D-LC method compatible with the MS method and protein extraction methods.

8.2) COLUMN PACKING

In order to develop functional columns for 1D and 2D-LC that could be made to any length or inside diameter or packed with any packing material, a column packing procedure was used for the packing of the columns used in our experiments. The setup involved a pressure chamber with a vial inside of it. Inside the vial was a packing material slurry, stirred by a magnetic stir-bar. One end of the capillary used to make the column was submerged in the vial, and the pressure inside the bomb was increased. The packing material from the slurry entered the capillary and traveled until it was stopped by a micro-filter or frit at the end of the capillary. Many arrangements were tried in the packing of columns. The type of filters/frits used to retain the packing material, the pressure used for column packing, the type of packing material, the solvents used to develop packing material slurries, the inside diameter and length of columns, and the

physical arrangement of the column packing procedure were all parameters that were adjusted to obtain ideal columns. Criteria used to assess the quality of a packed column were homogeneity, retention of analyte, and reproducibility in experiments.

8.3) LC METHODS

It was found that the speed of separation of bovine histones using an HPLC could be increased significantly by slightly altering the solvent gradient method used in separation. Since there were large gaps of time between peaks in the HPLC histone chromatograms, higher slopes of solvent percentage per time were used during these gaps. Therefore the elution time of many of the histones were decreased while the efficiency of the separation process was not compromised. The faster solvent gradient method that I developed is now in use in Dr. Michael Freitas' laboratory today.

8.4) CAPILLARY ELECTROPHORESIS

Many research hours were spent trying to interface a capillary electrophoresis instrument (CE) with an FT-ICR mass spectrometer via ESI to form the type of setup commonly referred to as CE-MS. There was no method for detection of analyte in the CE via UV or fluorescence, and therefore MS was the only possible method of detection with the available resources. In an improvised setup, I altered the electrical configuration of the CE to externalize much of the CE circuit and capillary and couple them with the ESI segment of the overall CE-MS setup. This project, though very educational, was a failure in terms of actual results.

8.5) ESI INTERFACING

During my research, I developed ESI tips and complex physical platforms for LC-MS interfacing. Factors that need to be considered in tip-making are the inside diameter at the tip, smoothness of the tip, and what method to use. The methods I used in tip making were pulling tips under flame, pulling tips in the presence of a very thin strip of ignited magnesium, and grinding tips with a diamond coated grinder. Tips made by the grinding method worked somewhat well, however were susceptible to instability of spray. These tips also did not have narrowed tips, which are necessary to mimic a point charges in ESI. Tips made by pulling capillaries over flame did have a narrowed diameter and therefore were better at mimicking point charges in ESI. However, the reproducibility obtained with these tips was poor due to variation in the result of the pulling process. Pulling tips in the presence of burning Mg was found to mimic a point charge better than either of the other methods in some cases and was found to create quickly tapered tips in some cases, because the presence of a thin sliver of very high temperature concentrated the heat in a small distance on the capillary. The Mg strip method, however, was not very reproducible due to the very high temperature involved.

In order to connect LC and CE techniques with MS via ESI, multiple physical ESI platforms were made. I designed platforms to conform to critical sizes required by connection locations on MS instruments and to conform to the necessary functions of holding the ESI and CE electrodes, holding columns and capillaries, and maintaining rigidity. The parts I drafted, using drawings and dimensions, were machined by the OSU Chemistry Dept. machine shop and are used in Dr. Michael Freitas' laboratory today.

9) REFERENCES

1) Background Information on the Human Proteome and the Yale/NHLBI Proteomics Center. (2002) Yale/NHLBI Proteomics Center. 1 pp. http://info.med.yale.edu/nhlbi-proteomics/proteome.htm
2) Zhang, L., Freitas, M. A., Wickman, J., Parthun, M. R., Klisovic, M. I., Marcucci, G., and Byrd, J. C. (2004) Differential Expression of Histone Post-translational Modifications in Acute Myeloid and Chronic Lymphocytic Leukemia Determined by High-pressure Liquid Chromatography and Mass Spectrometry. J. Am. Soc. Mass Spectrom. 15, 77-86.
3) Kimball, J. W. Red Blood Cells. Kimball's Biology Pages. (2006) 1 pp. http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/B/Blood.html#RBCs
4) Simpson, Richard J. Proteins and Proteomics. New York: Cold Spring Harbor Laboratory Press, 2003
5) Birktoft, J.J. and Breddam, K. (1994) In Proteolytic Enzymes: Serine and Cysteine Peptidases. A.J. Barrett (Eds.), Glutamyl endopeptidases. Methods Enzymol., 244, pp. 114-126. San Diego: Academic Press.
6) Drapeau, G.R., Boily, Y. and Houmard, J. (1972) Purification and properties of an extracellular protease of Staphylococcus aureus. J. Biol. Chem., 247, 6720-6726.
7) Cancer Deaths Decrease. Healthwatch. (Feb. 9, 2006) CBS Broadcasting. http://www.cbsnews.com/stories/2006/02/09/health/main1299899.shtml
8) Hoyert, Donna L., Heron, Melonie P., Murphy, Sherry L., Kung, Hsiang-Ching. (April 19, 2006) Deaths: Final Data for 2003. National Vital Statistics Reports. Volume 15, Number 13. http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_13.pdf
9) Berger, S. L. (2002) Histone modifications in transcriptional regulation. *Current Opinion in Genetics and Development* 12, 142-148.
10) Jenuwein, T., and Allis, C. D. (2001) Translating the histone code. *Science* 293, 1074-80.
11) Chirgadze, Dima. (2001) Introduction to Protein Crystallisation. Protein Crystallisation in Action. http://daffy.bioc.cam.ac.uk/~dima/whitepapers/xtal-in-action/node3.html
12) Marshall, Alan G. Alan Marshall's Homepage. Analytical Division, Department of Chemistry, Florida State University. http://www.chem.fsu.edu/faculty/marshall.htm
13) Turner Bryan, M. Cellular memory and the histone code. *CELL FIELD Publication Date:2002 Nov 1* 111, 285-91. FIELD Reference Number:38 FIELD Journal Code:0413066
14) Skoog, Holler, Nieman. Principles of Instrumental Analysis. Thompson Learning, 1998
15) Nägele, Edgar, Vollmer, Martin, Hörth, Patric, Vad, Cornelia. (2004) 2D-LC/MS Techniques for the Identification of Proteins in Highly Complex Mixtures. *Expert Rev. Proteomics*. 1(1) 37-46
16) Protein Synthesis. Access Excellence at the National Health Museum Resource Center. (2005) http://www.accessexcellence.org/RC/VL/GG/protein_synthesis.html
17) Cellular Differentiation. Wikipedia Online Encyclopedia. (May 12, 2006) http://en.wikipedia.org/wiki/Cellular_differentiation
18) Stem Cell Information. The Official National Institutes of Health Resource for Stem Cell Research. (August 12, 2005) http://stemcells.nih.gov/info/scireport/chapter1.asp
19) Nuyttens, Joachim. Rate Theory of Chromatography. (2006) http://openchemist.net/chemistry/chromatography/node9.php
20) Jonscher, Karen R., Yates, John R. The Whys and Wherefores of Quadrupole Ion Trap Mass Spectrometry. University of Washington at Seattle. (1996) http://www.abrf.org/ABRFNews/1996/September1996/sep96iontrap.html
21) Zhang, L., Freitas, M. A., Wickman, J., Parthun, M. R., Klisovic, M. I., Marcucci, G., and Byrd, J. C. (2004) Differential Expression of Histone Post-translational Modifications in Acute Myeloid and Chronic Lymphocytic Leukemia Determined by High-pressure Liquid Chromatography and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 15, 77-86.
22) Zhang, L., Eugeni, E. E., Parthun, M. R., and Freitas, M. A. (2003) Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* 112, 77-86.
23) El-Kom, Shibin. (1988) Functional and electrophoretic characteristics of faba bean (Vicia faba) flour proteins as affected by germination. National Center for Biotechnology Information. PMID: 3231250 (PubMed - indexed for MEDLINE)